

## Description of lncRNA transcript biotypes from Gencode release 12

([http://www.gencodegenes.org/gencode\\_biotypes.html](http://www.gencodegenes.org/gencode_biotypes.html))

**TEC:** To be Experimentally Confirmed. This is used for non-spliced EST clusters that have polyA features. This category has been specifically created for the ENCODE project to highlight regions that could indicate the presence of protein coding genes that require experimental validation, either by 5' RACE or RT-PCR to extend the transcripts, or by confirming expression of the putatively-encoded peptide with specific antibodies.

**retained\_intron:** Alternatively spliced transcript believed to contain intronic sequence relative to other, coding, variants.

**processed\_transcript:** Doesn't contain an ORF.

**non\_coding:** Transcript which is known from the literature to not be protein coding.

**ambiguous\_orf:** Transcript believed to be protein coding, but with more than one possible open reading frame.

**sense\_intronic:** Long non-coding transcript in introns of a coding gene that does not overlap any exons.

**sense\_overlapping:** Long non-coding transcript that contains a coding gene in its intron on the same strand.

**Antisense:** Transcript believed to be an antisense product used in the regulation of the gene to which it belongs.

**LincRNA:** Long, intervening noncoding (linc) RNAs, that can be found in evolutionarily conserved, intergenic regions.

**3prime\_overlapping\_ncrna:** Transcripts where ditag and/or published experimental data strongly supports the existence of short non-coding transcripts transcribed from the 3'UTR.